

pK_a measurements for the SAMPL6 prediction challenge for a set of kinase inhibitor-like fragments

Mehtap Işık^{1,2}, Dorothy Levorse³, Ariën S. Rustenburg^{1,4}, Ikenna Ndukwe⁵, Heather Wang⁵, David Mobley⁶, Timothy Rhodes³, John D. Chodera^{1*}

¹Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States; ²Tri-Institutional PhD Program in Chemical Biology, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065, United States; ³Merck & Co., Inc., MRL, Pharmaceutical Sciences, 126 East Lincoln Avenue, Rahway, New Jersey 07065, United States; ⁴Graduate Program in Physiology, Biophysics, and Systems Biology, Weill Cornell Medical College, New York, NY 10065, United States; ⁵Merck Co., Inc., MRL, Process Research & Development, 126 East Lincoln Avenue, Rahway, New Jersey 07065, United States; ⁶Department of Pharmaceutical Sciences and Department of Chemistry, University of California, Irvine, Irvine, California 92697, United States

***For correspondence:**

john.chodera@choderalab.org (JDC)

Abstract Determining the protonation state of a small molecule is a preliminary requirement for predicting its physicochemical and pharmaceutical properties, as well as interactions with protein targets using computational models. To determine the ionic state of a molecule in an aqueous solution at a certain pH it is necessary to know its acid dissociation constants (pK_a values). As a part of SAMPL6 community challenge, we organized a blind pK_a prediction component to assess the accuracy of contemporary pK_a prediction methods. While inaccuracy in prediction of small molecule pK_as have potential detrimental impact on predictive physical models, predicting pK_as of drug-like molecules can be difficult due to challenging properties such as multiple titratable sites, heterocycles, and tautomerization. We limited the focus of this challenge on a subset of chemical space of drug-like molecules: 24 small molecules were selected to represent fragments of kinase inhibitors. We measured macroscopic pK_a values of SAMPL6 compounds with UV-absorbance based method with Sirius T3 instrument to construct an experimental reference dataset for the evaluation of computational pK_a predictions.

Keywords

acid dissociation constants · spectrophotometric pK_a measurement · blind prediction challenge · SAMPL · macroscopic pK_a · microscopic pK_a · macroscopic protonation state · microscopic protonation state

Abbreviations

SAMPL Statistical Assessment of the Modeling of Proteins and Ligands

pK_a $-\log_{10}$ acid dissociation equilibrium constant

psK_a $-\log_{10}$ apparent acid dissociation equilibrium constant in cosolvent

DMSO Dimethyl sulfoxide

ISA Ionic-strength adjusted
SEM Standard error of the mean
TFA Target factor analysis
LC-MS Liquid chromatography - mass spectrometry

Introduction

Brief summary, origin and goal of SAMPL6 and outlook for future challenges.

Selection criteria of SAMPL6 compounds

Overview of the concept of microscopic and macroscopic pKas.

Overview of pKa measurement methods.

Spectrophotometric pKa method is more sensitive than potentiometric method and requires very low analyte concentrations (~ 50 μ M) which is advantageous especially for compounds with low solubility, but it is only applicable to titration sites near chromophores. For protonation to effect absorbance protonation site should be a maximum of 4 heavy atoms away from the chromophore: conjugated double bonds, carbonyl groups, aromatic rings etc. Although potentiometric measurement doesn't have this structural limitation, a higher concentration of analyte (~ 5 mM) is necessary for potentiometric method than for spectrophotometric method to provide large enough buffering capacity signal above water for an accurate measurement. Therefore, we have decided to use spectrophotometric measurements for collecting experimental pKa data and selected SAMPL6 pKa challenge compounds constructing SAMPL6 pKa dataset and se for pKa measurements

Brief outline of what this paper includes.

Methods

Compound selection and procurement

Selection criteria of SAMPL6 compounds

Small molecules were purchased in powder form.

FIGURE for compound filtering steps. Flow chart or funnel for Illustration of compound selection and experimental measurement workflow.

SI TABLE vendor, lot numbers, SMILES. Selected compounds and their SAMPL challenge identities, eMolecules ID, supplier reported purity, LC/MS purity

UV-metric pKa measurements

Experimental pKa measurements were collected using spectrophotometric pKa method with a Sirius T3 automated titrator instrument (Pion) at 25°C and constant ionic strength. Sirius T3 instrument is equipped with an Ag/AgCl double-junction reference electrode to monitor pH, a dip probe attached to spectrophotometer that measures, a stirrer, and automated volumetric titration capability. The UV-metric pKa measurement protocol of the Sirius T3 measures the change in multiwavelength absorbance in the 250-450 nm UV region of the absorbance spectrum while the pH is titrated between pH 1.8 and 12.2 to evaluate pKas [1, 2].

DMSO stock solutions of each compound with 10 mg/ml concentration were prepared by weighing 1 mg of powder chemical with Sartorius Analytical Balance (Model: ME235P) and dissolving it in 100 μ L DMSO (Dimethyl sulfoxide, Fisher Bioreagents, CAT: BP231-100, LOT: 116070, purity \geq 99.7%). DMSO stock solutions were capped immediately to limit hygroscopicity of DMSO and sonicated for 5-10 minutes in water bath sonicator at room temperature to ensure proper dissolution. These DMSO stock solutions were stored in room temperature up to 2 weeks. 10 mg/ml DMSO solutions were used as stock solutions for the preparation of 3 replicate samples for independent titrations: 15 μ L of 10 mg/ml DMSO stock solution delivered to 4 mL glass sample vials of Sirius T3 with an electronic micropipette (Rainin EDP3 LTS 1-10 μ L). The volume of

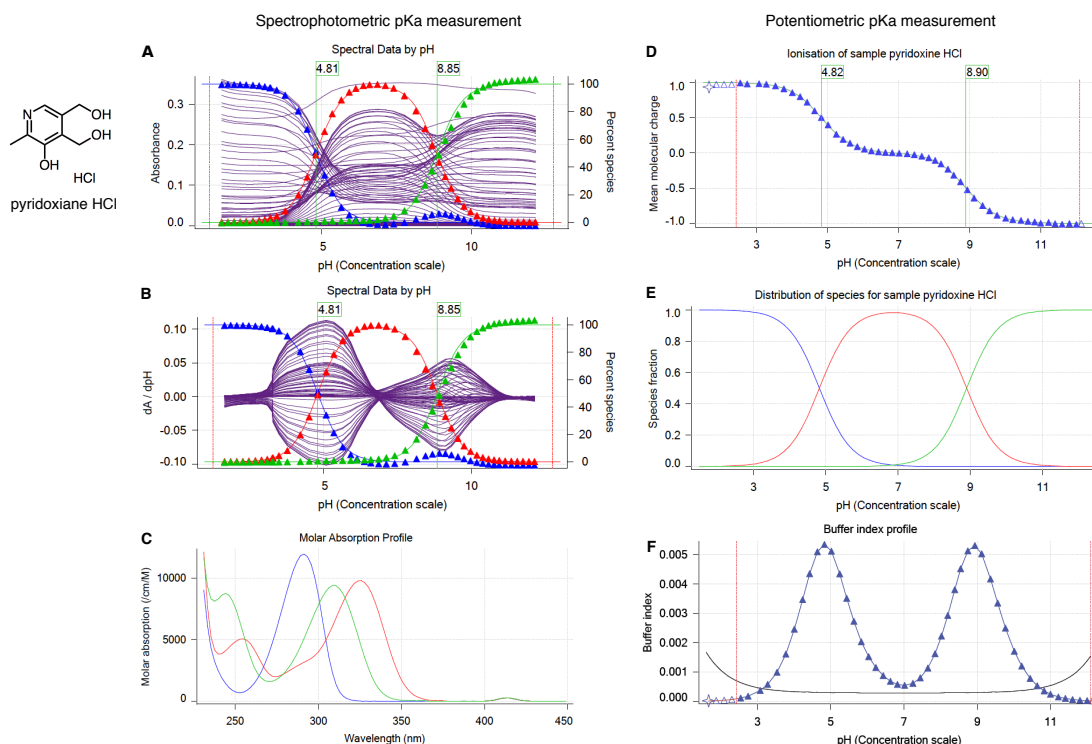


Figure 1. Spectrophotometric (UV-metric) and potentiometric (pH-metric) pKa measurements of pyridoxine HCl with Sirius T3. Spectrophotometric method (panels **A**, **B**, **C**) relies on UV-absorbance difference between protonation states. It is very sensitive and requires very low analyte concentrations (~ 50 μ M), but limited to titration sites near chromophores. **A** Multiwavelength absorbance vs pH. Purple lines represents absorbance at each wavelength in UV-region. **B** Derivative of multiwavelength absorbance with respect to pH (dA/dpH) vs pH is plotted with purple lines. In **A** and **B** Blue, red, and green triangles represent population of protonation states (from most protonated to least protonated) as calculated from experimental data at each pH value. pKa values (green flags) correspond to inflection point of multiwavelength absorbance data where change in absorbance with respect to pH is maximum. **C** Molar absorption coefficients vs wavelength for each protonation state as resolved by TFA. **D**, **E**, **F** illustrate potentiometric pKa measurement where molar addition of acid or base is tracked as pH is titrated. **D** Mean molecular charge vs pH. Mean molecular charge is calculated based on the model provided for the analyte: predicted number and nature of titratable sites (acid or base type), and number of counter ions present. pKa values are calculated as inflection points of charge vs pH plot. **E** Predicted macroscopic protonation state populations vs pH calculated based on pKa values (H₂A⁺: blue, HA: red, and A⁻: green) **F** Buffering index vs pH profile of water (grey solid line, theoretical) and the sample solution (blue triangles represent experimental data points). A higher concentration of analyte (~ 5 mM) is necessary for potentiometric method than for spectrophotometric method to provide large enough buffering capacity signal above water for an accurate measurement.

delivered DMSO stock solution, which determines the sample concentration, is optimized individually for each compound to achieve sufficient but not saturated absorbance signal (targeting 0.5-1.0 AU) in the linear response region. Another limiting factor for sample concentration was ensuring that the compounds stays soluble in the whole range of pH titration range. 25 μ L of mid-range buffer (14.7 mM K_2HPO_4 0.15 M KCl in H_2O) was added to each sample transferred with a micropipette (Rainin EDP3 LTS 10-100 μ L) to provide enough buffering capacity in middle pH ranges so that pH could be controlled incrementally throughout the titration.

pH is temperature and ionic-strength dependent. Sample heating block of Sirius T3 kept the analyte solution at 25 ± 0.5 °C throughout the titration. Ionic-strength of the samples were adjusted by dilution in 1.5 mL ionic-strength adjusted water (ISA water, 0.15 M KCl) to the vials by Sirius T3. Analyte dilution, mixing, acid/base titration, and measurement of UV-absorbance was automated by UV-metric pKa measurement protocol of Sirius T3(Pion). pH was titrated between pH 1.8 and 12.2 with addition of acid (0.5 M HCl) and base (0.5 M KOH) targeting 0.2 pH steps between datapoints. Titrations were performed under Argon flow on the surface of sample solution to limit carbondioxide absorption from air. To capture all sources of experimental variability fully, instead of doing 3 tandem pH titrations on the same sample solution, pKas of three replicate samples were measured separately with one round of pH titration. Although this choice has higher cost for throughput and analyte consumption, it limits the dilution of the analyte during multiple titrations and drop in absorbance signal.

Visual inspection of sample solutions after titration and inspection of pH-dependent absorbance shift in 500-600 nm region of UV spectra was used to verify no detectable precipitation occurred during the course of measurement. Absorbance in 500-600 nm region of UV spectra is associated with scattering of longer wavelengths of light in the presence of aggregates. For each analyte, we optimized analyte concentration, direction of titration and the range of pH titration in order to achieve solubility. The direction of titration was determined so that titration would start from the pH where the compound is most soluble: low-to-high pH for bases and high-to-low pH for acids. UV-metric pKa measurement method typically requires as low as 50 μ M sample concentration (although minimum concentration requirement may change based on absorbance properties of the analyte). Some compounds may not be soluble even at such low concentration throughout pH range of the titration. As the sample is titrated through a wide range of pH values, it is likely that low solubility ionization states such as neutral and zwitterionic states will be also be visited, limiting the highest analyte concentration that can be titrated without encountering any solubility issues. For compounds with insufficient solubility to accurately determine a pKa directly in a UV-metric titration, a cosolvent protocol was used [See the next section: UV-metric pKa measurement with cosolvent].

Two Sirius T3 softwares, Sirius T3Control v1.1.3.0 and Sirius T3Refine v1.1.3.0, were used to execute measurement protocols and to analyze pH-dependent multivelength spectra, respectively. A protonation state change of titratable sites near chromophores will modulate the UV-absorbance spectra of these chromophores, allowing populations of distinct UV-active species to be resolved as a function of pH. To do this, basis spectra are identified and populations extracted via TFA analysis of the pH-dependent multi-wavelength absorbance [2]. When fitting the absorbance data to a titratable molecule model to estimate pKas, we selected the minimum number of pKas that was sufficient to provide a high quality of fit between experimental and modeled data based on visual inspection of pH-dependent populations.

This method is capable of measuring pKas between 2 and 12 when protonatable groups are at most 4-5 heavy atoms away from chromophores such that a change in protonation state alters the absorbance spectrum of the chromophore. We have selected compounds where titratable groups are close to potential chromophores (generally aromatic ring systems), but it is possible that our experimental results couldn't detect protonation of titratable groups distal to a UV-chromophore.

UV-metric pKa measurement with cosolvent

If analytes are not sufficiently soluble in aqueous environment, pKa values can not be determined accurately with UV-metric pKa measurements method. If precipitation is present, the UV-absorbance signal from pH-dependent precipitate formation can not be differentiated from the pH-dependent protonation signal. For compounds with low aqueous solubility pKa values were estimated from multiple apparent pKas

measurements done in methanol:water cosolvent solutions with various ratios. This method is referred to as UV-metric pKa method in Sirius T3 Manual [3].

In cosolvent spectrophotometric pKa measurement protocol was very similar to UV-metric pKa measurement method in water except the following aspects: Three titrations were performed in typically in 30%, 40%, and 50% mixtures methanol:water by volume to measure apparent pKa values (psKa) in these mixtures. Then Yasuda-Shedlovsky extrapolation method was used to estimate the pKa value at 0% cosolvent [4–6].

$$psKa + \log[H_2O] = A/\epsilon + B$$

Yasuda-Shedlovsky extrapolation relies on the linear correlation between $psKa + \log[H_2O]$ and reciprocal dielectric constant of the cosolvent mixture ($1/\epsilon$). In the equation above, A and B are the slope and intercept of the line fitted to experimental datapoints. Depending on the solubility requirements of the analyte methanol ratio of the cosolvent mixtures were adjusted. We designed the experiments to have at least 5% cosolvent ratio difference between datapoints and 60% methanol at most. Calculation of Yasuda-Shedlovsky extrapolation was performed by Sirius T3 software using at least 3 psKa values measured in different ratios of methanol:water. Addition of methanol (ISA, 0.15 M KCl) was controlled by the instrument before each titration. Three consecutive pH titrations at different methanol concentrations were performed using the same sample solution. In addition, three replicate measurements with independent samples (prepared from the same DMSO stock) were collected.

Calculation of uncertainty in pKa measurements

Experimental uncertainties were reported as standard error of the mean (SEM) of three replicate pKa measurements. Since Sirius T3 software reports pKa values with 2 decimal places, we have reported SEM as 0.01 in cases where SEM values calculated from 3 replicates were lower than 0.01.

Protonation site determination with NMR

Update or remove this section based on updates from lyke.

Quality control for chemicals

Purity of SAMPL6 pKa challenge compounds were determined based on LC-MS. The purity analysis was performed using an Agilent HPLC1200 Series equipped with auto-sampler, UV diode array detector and a Quadrupole MS detector 6140. The software used was Chemstation for LC & LCMS with version C01.07SR2. The column for the analysis is Assentis Express C18 3.0x100mm 2.7 µl particle size at Column temperature = 45°C.

- Mobile phase A: 2 mM ammonium formate (pH=3.5) aqueous
- Mobile phase B: 2 mM ammonium formate in Acetonitrile : Water=90:10 (pH=3.5)
- Flow rate : 0.75ml/min
- Gradient: Starting with 10% B to 95%B in 10 minutes then hold 95%B for 5 minutes.
- Post run length: 5 minutes
- Mass condition: ESI positive and negative mode
- Capillary voltage: 3000 V
- Drying gas flow: 12 ml/min
- Nebulizer pressure: 35 psi
- Drying temperature: 350°C
- Mass range: 5-1350 Da; Fraggmentor:70; Threshold:100

The percent area for main peak is calculated based on the area of main peak divided by the total area of all peaks. The percent area of the main peak is reported as an estimate of sample purity.

SI Figure: all MS data

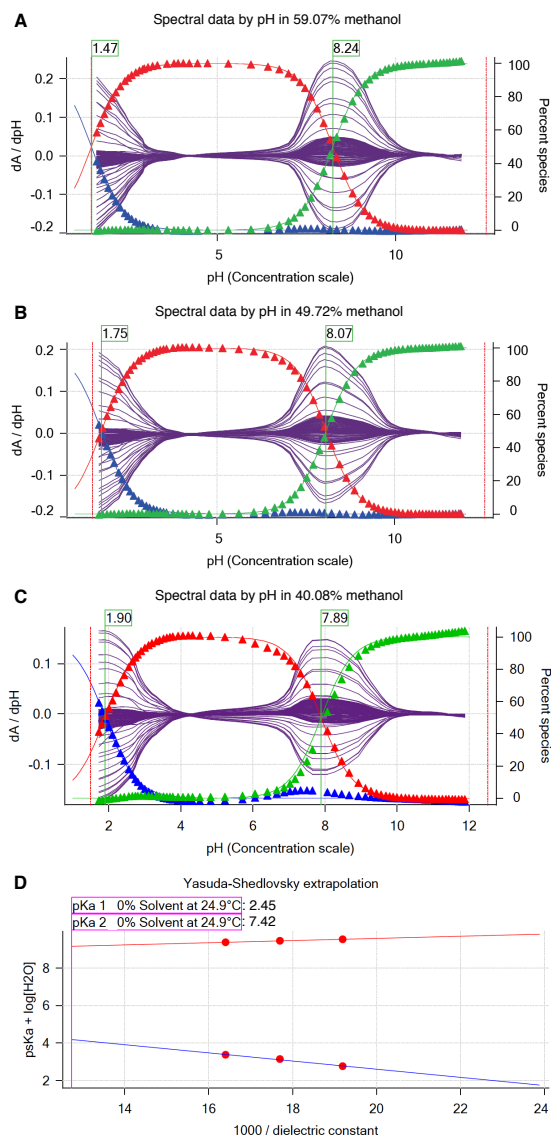


Figure 2. Determination of SM22 pKa values with cosolvent method and Yasuda-Shedlovsky extrapolation. **A**, **B**, and **C** show $psKa$ of SM22 determined at various methanol concentrations: 59.07%, 49.72%, 40.08% by weight. Purple solid lines indicate derivative of absorbance with respect to pH vs pH in multiple wavelength. $psKa$ values (green flags) were determined by Sirius T3 Refinement Software. Blue, red and green triangles show relative populations of macroscopic protonation states with respect to pH calculated from the experimental data. Notice that as cosolvent concentration increases, $psKa$ 1 value decreases from 1.90 to 1.47 and $psKa$ 2 value increases from 7.84 to 8.24. **D** Yasuda-Shedlovsky extrapolation plot for SM22. Red datapoints correspond to $psKa$ determined at various cosolvent ratios. Based on linear fitting to $psKa + \log[H_2O]$ vs $1/\epsilon$, pKa 1 and pKa 2 in 0% cosolvent (aqueous solution) was determined as 2.45 and 7.42, respectively. R^2 values of linear fits are both 0.99. The slope of Yasuda-Shedlovsky extrapolation shows if the observed titration has acidic(positive slope) or basic(negative) character dominantly, although this is an macroscopic observation and should not be relied on for annotation of $pKas$ to functional groups (microscopic $pKas$).

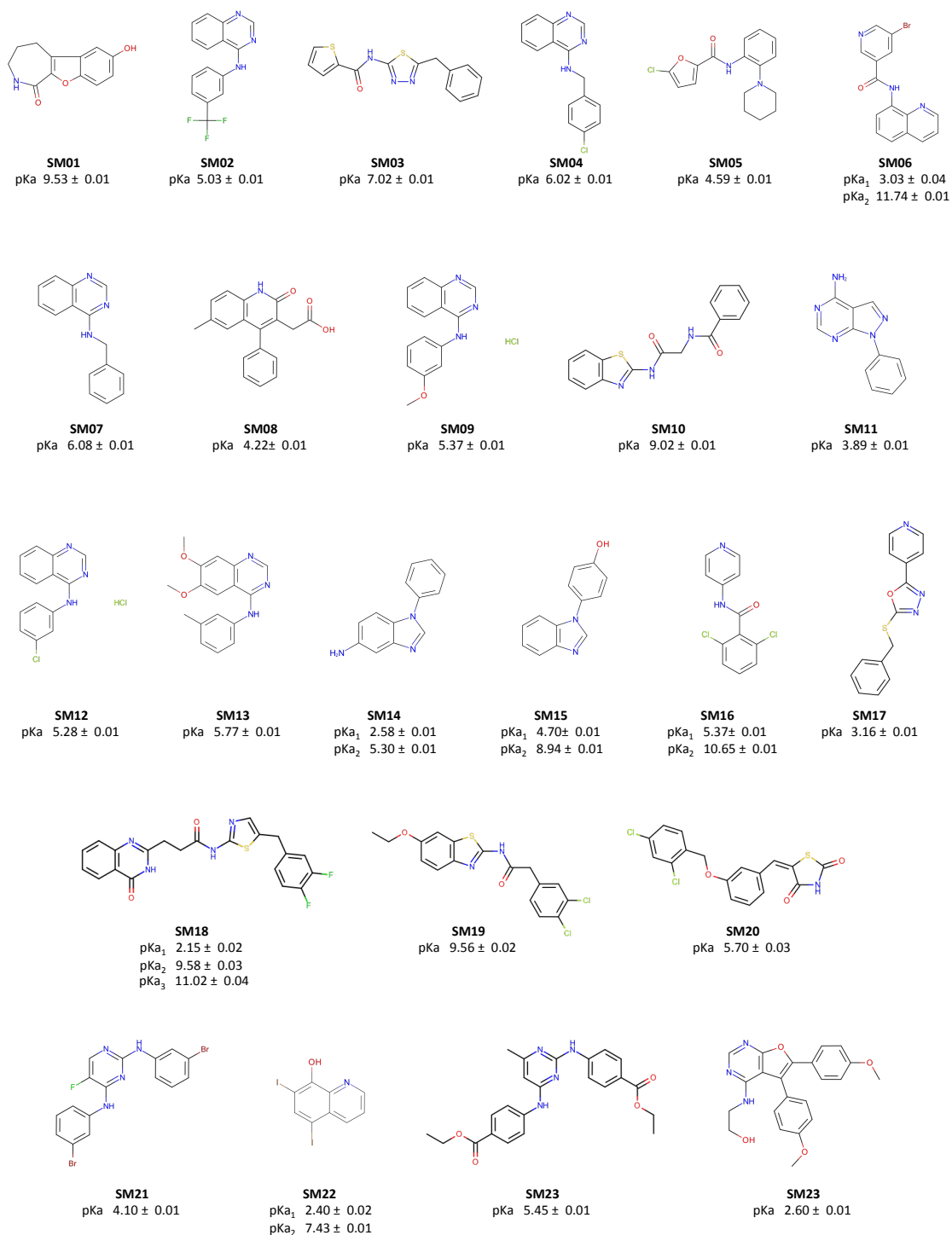


Figure 3. Molecules used in the SAMPL6 pKachallenge. Experimental UV-metric pKameasurements were performed for these 24 molecules, and discernable macroscopic pKas reported. Uncertainties are expressed as the standard error of the mean of three independent measurements.

Results

TABLE: Measured macroscopic pKa for all compounds, what method

Effect of cosolvent on measured pKas

FIGURE: Cosolvent vs non-cosolvent measurement difference

Discussion

Discussion of experimental data interpretation: macroscopic

Multiwavelength absorbance analysis on the Sirius T3 allows for very good resolution of pKas, but it is important to note that this method produces estimates of macroscopic pKas. If multiple microscopic pKas have close pKa values and overlapping changes in UV absorbance spectra associated with protonation/deprotonation event, the spectral analysis could produce a single macroscopic pKa that represents an aggregation of multiple microscopic pKas.

Warning about acid base assignment from cosolvent measurements

Lessons learned for future challenge iterations

Challenges with NMR measurements?

pKa can shift based on ionic strength, temperature, and lipophilic content Can other cosolvents be interesting for process chemistry.

Code and data availability

- SAMPL6 pKa challenge instructions, submissions, and analysis is available at <https://github.com/MobleyLab/SAMPL6>
- Python scripts used for compound selection are available at **compound_selection** directory of <https://github.com/choderalab/sampl6-physicochemical-properties>

Construct a proper README for compound selection directory.

Overview of supplementary information

Author Contributions

Complete this section.

Conceptualization, ; Methodology, ; Software, ; Formal Analysis, ; Investigation, MI, ; Resources, ; Data Curation, ; Writing-Original Draft, ; Writing - Review and Editing, ; Visualization, ; Supervision, ; Project Administration, ; Funding Acquisition,

Acknowledgments

Complete this section.

MI, ASR, and JDC acknowledge support from the Sloan Kettering Institute. JDC acknowledges support from NIH grant P30 CA008748. MI acknowledges Doris J. Hutchinson Fellowship. Acknowledging Merck Preformulation department for materials, expertise and instrument time.

Brad Sherborne Paul Czodrowski - helped with purchasable compound list in earlier iteration Ikenna Ndukwe Caitlin Bannan

Disclosures

JDC is a member of the Scientific Advisory Board for Schrödinger, LLC.

References

- [1] **Tam KY**, Takács-Novák K. Multi-wavelength spectrophotometric determination of acid dissociation constants: a validation study. *Analytica chimica acta*. 2001; 434(1):157–167.
- [2] **Allen RI**, Box KJ, Comer JEA, Peake C, Tam KY. Multiwavelength spectrophotometric determination of acid dissociation constants of ionizable drugs. *Journal of pharmaceutical and biomedical analysis*. 1998; 17(4):699–712.
- [3] Sirius T3 User Manual, v1.1. Sirius Analytical Instruments Ltd, East Sussex, UK; 2008.
- [4] **Avdeef A**, Box KJ, Comer JEA, Gilges M, Hadley M, Hibbert C, Patterson W, Tam KY. PH-metric logP 11. pK a determination of water-insoluble drugs in organic solvent–water mixtures. *Journal of pharmaceutical and biomedical analysis*. 1999; 20(4):631–641.
- [5] **Avdeef A**, Comer JEA, Thomson SJ. pH-Metric log P. 3. Glass electrode calibration in methanol-water, applied to pKa determination of water-insoluble substances. *Analytical Chemistry*. 1993; 65(1):42–49. <https://doi.org/10.1021/ac00049a010>, doi: 10.1021/ac00049a010.
- [6] **Takács-Novák K**, Box KJ, Avdeef A. Potentiometric pKa determination of water-insoluble compounds: validation study in methanol/water mixtures. *International Journal of Pharmaceutics*. 1997; 151(2):235 – 248. <http://www.sciencedirect.com/science/article/pii/S0378517397049077>, doi: [https://doi.org/10.1016/S0378-5173\(97\)04907-7](https://doi.org/10.1016/S0378-5173(97)04907-7).
- [7] **Manallack DT**. The pK(a) Distribution of Drugs: Application to Drug Discovery. *Perspectives in Medicinal Chemistry*. 2007 Sep; 1:25–38.

222 **Supplementary Information**